Supporting Information S1. Reasons for excluding certain complexes from the study.

Reasons for exclusion	Complex
Protein binding sites are incomplete: the ligand in each of these complexes interacts with a dimer or a tetramer of the protein. However, only part of the dimer or tetramer is given in the database.	6FIV, 1B11, 1TYR, 2D1O, 2USN
Ligand has less than three pharmacophores therefore cannot be matched onto the protein pharmacophores in clique detection.	1L83
Large protein-ligand complexes that need more than four hours to finish: Both the ligand and the number of protein pharmacophores generated for these complexes are typically quite large. It is too time-consuming to perform an exhaustive search by clique detection on these systems. We are developing a new method to treat these large protein-ligand complexes. Therefore no extra effort was spent in handling these complexes in this study.	1FKN, 1FO0, 1FZK, 1G7Q, 1HFS, 1JQ9, 1NNY, 1PZ5, 1SLG, 1XD1, 2B7D, 2BZZ, 2ER9, 4FIV
Total	20

Supporting Information S2. Potential functions for computing interaction energies on the 3D grid.

The interaction potentials for hydrogen-bonding and hydrophobic ligand atoms placed at individual grid points were computed using a continuous form of the ChemScore[1-2] scoring function. The aromatic and ionic interactions were calculated using a functional form similar to ChemScore. In detail, the interaction potential for a hydrogen-bond donating atom *i* on a grid point *j* was computed by

$$V_j^{donor} = V_{j,dist}^{donor} \cdot V_{j,angle}^{donor} \tag{1a}$$

with

$$V_{j,dist}^{donor} = \begin{cases} 0 & r_{ij} > 3.4\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.5\ \text{Å}} \cdot \left(r_{ij} - r_{ij}^{HB} - 0.2\ \text{Å}\right)\right) & 2.9\ \text{Å} < r_{ij} \leq 3.4\ \text{Å} \\ -1 & 2.5\ \text{Å} < r_{ij} \leq 2.9\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.5\ \text{Å}} \cdot \left(-r_{ij} + r_{ij}^{HB} - 0.2\ \text{Å}\right)\right) & 2.0\ \text{Å} < r_{ij} \leq 2.5\ \text{Å} \\ 0 & r_{ij} \leq 2.0\ \text{Å} \end{cases} \tag{1b}$$

and

$$V_{j,angle}^{donor} = \begin{cases} 0 & \cos\varphi < \cos(88°) \\ 0.5 + 0.5 \cdot \cos\left(\frac{\pi}{\cos(88°) - \cos(27°)} \cdot (\cos\varphi - \cos(27°))\right) & \cos(88°) < \cos\varphi < \cos(27°) \\ 1 & \cos\varphi > \cos(27°) \end{cases} \tag{1c}$$

Where r_{ij} was the distance between the heavy atom *i* and grid point *j*. r_{ij}^{HB} is the approximated sum of van der Waals radii of the donor heavy atom and hypothetical acceptor atom placed on the grid. The angle φ was defined by the angle between lone pair, acceptor of protein and grid point. The same functional form was used for hydrogen-bond acceptors with φ defined by the angle between protein's donor hydrogen atom, donor heavy atom and grid point. The hydrophobic potential was computed by

$$V_j^{hydrophobic} = \begin{cases} 0 & r_{ij} > r_{ij}^{vdW} + 1.7\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{1.4\ \text{Å}} \cdot \left(r_{ij} - r_{ij}^{vdW} - 0.3\ \text{Å}\right)\right) & r_{ij}^{vdW} + 0.3\ \text{Å} < r_{ij} \leq r_{ij}^{vdW} + 1.7\ \text{Å} \\ -1 & 2.0\ \text{Å} < r_{ij} \leq r_{ij}^{vdW} + 0.3\ \text{Å} \\ 0 & r_{ij} \leq 2.0\text{Å} \end{cases}$$

(2)

r_{ij}^{vdW} is the sum of van der Waals radii of protein atom and grid point. The grid point can be considered to represent a potential binding position of a hydrophobic ligand atom. Thus, the van der Waals radius of a carbon atom is assumed for the grid point as carbon atoms are most frequently engaged in hydrophobic contacts between protein and ligand. The distance and angle threshold values in equations 1 and 2 were adjusted to reproduce the overall form of the original ChemScore scoring function. The modified function, however, provides continuous derivatives of the potential with respect to the coordinates.

For aromatic interactions, both sandwich (AromP) and T-shaped (AromT) configurations were considered for a given aromatic group:

$$V_j^{aromP} = V_{j,dist}^{aromP} \cdot V_{j,angle}^{arom} \quad \text{and} \quad V_j^{aromT} = V_{j,dist}^{aromT} \cdot V_{j,angle}^{arom} \tag{3a}$$

with

$$V_{j,dist}^{aromP} = \begin{cases} 0 & r_{ij} > 4.3\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.6\ \text{Å}} \cdot \left(r_{ij} - 3.7\ \text{Å}\right)\right) & 3.7\ \text{Å} < r_{ij} \leq 4.3\ \text{Å} \\ -1 & 3.5\ \text{Å} < r_{ij} \leq 3.7\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.6\ \text{Å}} \cdot \left(3.5 - r_{ij}\right)\right) & 2.9\ \text{Å} < r_{ij} \leq 3.5\ \text{Å} \\ 0 & r_{ij} \leq 2.9\ \text{Å} \end{cases} \tag{3b}$$

$$V_{j,dist}^{aromT} = \begin{cases} 0 & r_{ij} > 6.1\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.7\ \text{Å}} \cdot (r_{ij} - 5.4\ \text{Å})\right) & 5.4\ \text{Å} < r_{ij} \leq 6.1\ \text{Å} \\ -1 & 4.9\ \text{Å} < r_{ij} \leq 5.4\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.7\ \text{Å}} \cdot (4.9 - r_{ij})\right) & 4.2\ \text{Å} < r_{ij} \leq 4.9\ \text{Å} \\ 0 & r_{ij} \leq 4.2\ \text{Å} \end{cases} \quad (3c)$$

and

$$V_{j,angle}^{donor} = \begin{cases} 0 & \cos\varphi \leq \cos(45°) \\ 0.5 + 0.5 \cdot \cos\left(\frac{\pi}{\cos(45°) - \cos(30°)} \cdot (\cos\varphi - \cos(30°))\right) & \cos(30°) < \cos\varphi \leq \cos(45°) \\ 1 & \cos\varphi > \cos(30°) \end{cases}$$

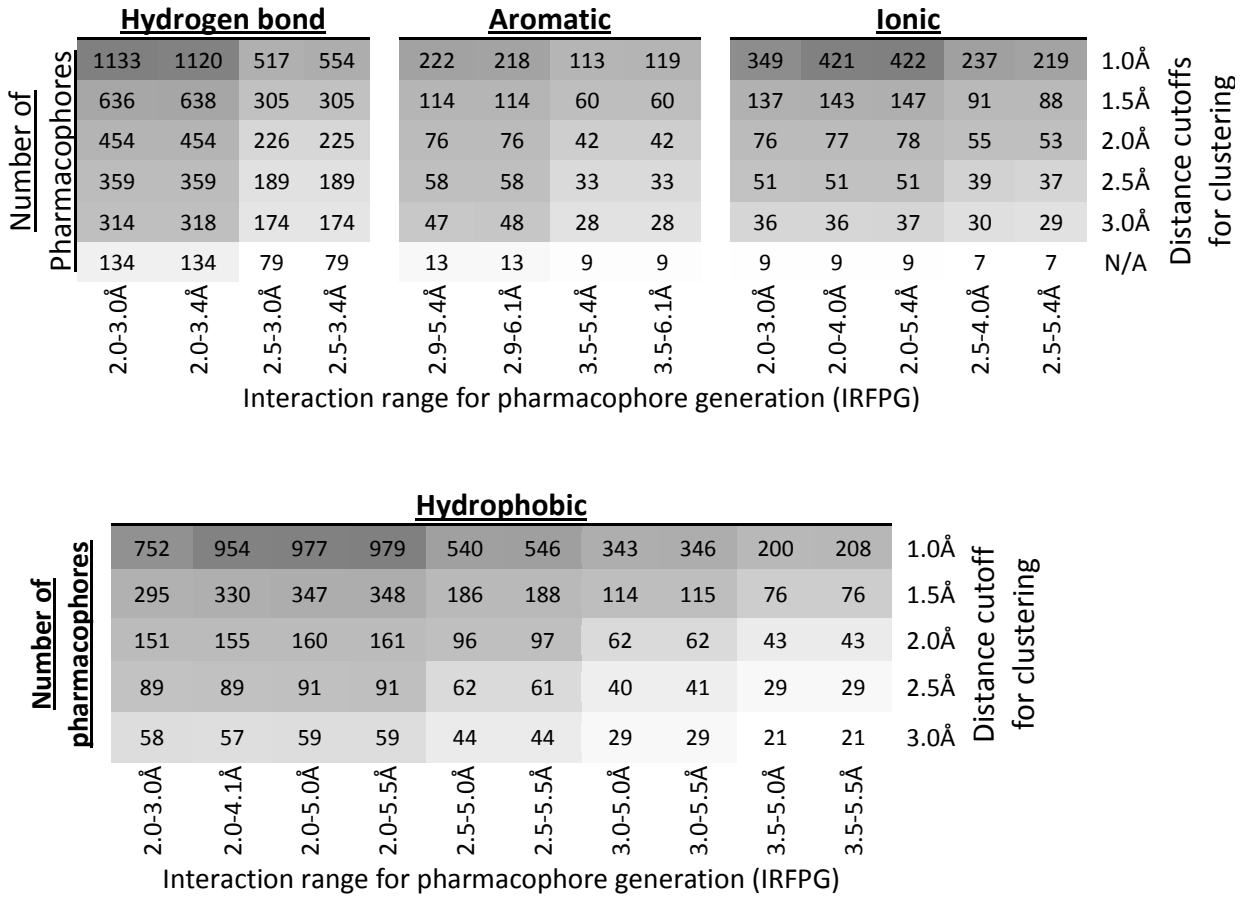
(3d)

Finally, the ionic interactions were computed by:

$$V_{j,dist}^{ionic} = \begin{cases} 0 & r_{ij} > 5.4\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.5\ \text{Å}} \cdot (r_{ij} - 2.9\ \text{Å})\right) & 2.9\ \text{Å} < r_{ij} \leq 5.4\ \text{Å} \\ -1 & 2.5\ \text{Å} < r_{ij} \leq 2.9\ \text{Å} \\ -0.5 - 0.5 \cdot \cos\left(\frac{\pi}{0.5\ \text{Å}} \cdot (-r_{ij} + 2.5\ \text{Å})\right) & 2.0\ \text{Å} < r_{ij} \leq 2.5\ \text{Å} \\ 0 & r_{ij} \leq 2.0\ \text{Å} \end{cases} \quad (4)$$

Where r_{ij} was the distance between the heavy atom *i* and grid point *j* with opposite ionic properties.

Supporting Information S3. Average number of protein pharmacophores per protein-ligand complex generated using different set of parameters.



References

1. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997,** *11* (5), 425-45.

2. Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D., Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins: Struct., Funct., Bioinf.* **1998,** *33* (3), 367-382.